DOCUMENT RESUME

ED 400 301 TM 025 619

| | |
|---|---|
| AUTHOR | Ligon, Glynn D. |
| TITLE | Data Quality: Earning the Confidence of Decision Makers. |
| PUB DATE | Apr 96 |
| NOTE | 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996). |
| PUB TYPE | Reports - Descriptive (141) -- Speeches/Conference Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | Classification; Data Collection; *Decision Making; *Educational Research; *Evaluation Methods; *Quality Control; *User Needs (Information) |
| IDENTIFIERS | *Data Quality |

ABSTRACT

Professionals responsible for educational research, evaluation, and statistics have sought to provide timely and useful information to decision makers. Regardless of the evaluation model, research design, or statistical methodology employed, informing the decision making process with quality, reliable data is a basic goal. The definition of quality for education data has not been adequately addressed in the literature of educational research and evaluation. In the publications describing quality related to general information systems, the concept is narrowly interpreted to mean accurately and reliably processed data. This paper ties together the foundations of data quality from the formal information systems literature with the practical data quality in the arena of public education decision making. A hierarchy of data quality is described to assist both the understanding of quality and the requirements for achieving quality. The hierarchy ranges from the availability of dysfunctional, bad data to the quality level of data-based decisions made with confidence. For practitioners, a checklist is provided for use in determining the quality of their data sources. Attachments include the data quality typology, a list of ratings of data quality, and the checklist for rating and improving data quality. (Author/SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

# Data Quality: Earning the Confidence of Decision Makers

*Glynn D. Ligon*
*Evaluation Software Publishing, Incorporated*
*Paper Presented at the Annual Meeting of the American Educational Research Association*
*April 1996 New York, New York*

ED 400 301

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
GLYNN D. LIGON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

**Data quality is more than accuracy and reliability. High levels of data quality are achieved when information is valid for the use to which it is applied, and when decision makers have confidence in the data and rely upon them.**

## Summary

Professionals responsible for educational research, evaluation, and statistics have sought to provide timely and useful information to decision makers. Regardless of the evaluation model, research design, or statistical methodology employed, informing the decision making process with quality, reliable data is a basic goal. The definition of quality for education data has not been adequately addressed in the literature of educational research and evaluation. In the publications describing quality related to general information systems, the concept is narrowly interpreted to mean accurately and reliably processed data. This paper ties together the foundations of data quality from the formal information systems literature with the practical aspects of data quality in the arena of public education decision making. A hierarchy of data quality is described to assist both the understanding of quality and the requirements for achieving quality. The hierarchy ranges from the availability of dysfunctional, bad data to the quality level of data-based decisions made with confidence. For practitioners, a checklist is provided for use in determining the quality of their data sources.

Readers of this paper are requested to provide the author with ideas on the topic of data quality. Comments specific to this paper, anecdotes illustrating points, or further thinking related to the pursuit of data quality are all solicited. Please communicate your reactions to:

| | |
|---|---|
| Internet: | gligon@evalsoft.com |
| Voice: | 512-458-8364 |
| Fax: | 512-371-0520 |
| Mail: | 3405 Glenview Avenue Austin, Texas 78703 |

Copies of the paper may be downloaded from ESP's FTP server as follows:

challenger.tpoint.net
(Use anonymous login; change directory to /evalsoft/swstand.)

## Background

Data quality is essential to successful research, evaluation, and statistical efforts in public schools. As statewide accountability systems that rely upon large data bases grow, concern follows about the data quality within those emerging state-level data bases. As states and the Federal government move toward establishing data warehouses to make information available electronically to anyone, questions are raised about the quality of the data collected and stored. What is not universally sought is Federally imposed standards for data and information systems. There is broad support for voluntary standards which states and local school districts can adopt.



BEST COPY AVAILABLE

TM025619

What is needed first is a way to know when quality data are available and when caution should be exercised. (***New Developments in Technology: Implications for Collecting, Storing, Retrieving, and Disseminating National Data for Education*** G. Ligon, Paper Prepared for MPR Associates and the National Center for Education Statistics, November, 1995.)

Decision makers at all levels are relying upon data to inform, justify, and defend their positions on important issues. What are the key criteria on which to determine data quality? Is there a logical sequence to the processes for ensuring quality in information systems?

The concern for data quality is somewhat different than the slowly emerging interest in education data that has grown for decades. The concern for data quality is a sign of maturity in the field, an increasing sophistication by the audiences who use education data. In other words, first we asked "Are our students learning?" Then we had to ask "What are the education indicators that we should be monitoring?" Finally, we are asking "Now that we have some indicators, do we trust them?" (***What Dow-Jones Can Teach Us: Standardized Education Statistics and Indicators,*** G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1993; ***A Dow Jones Index for Educators***, G. Ligon, The School Administrator, December, 1993.)

An easy point in time to mark is the release of the "Nation at Risk" report. Much reform in education followed, including expansion of accountability systems within states. The search heated up for the true, reliable indicators of quality in education. A major event was the passage of the 1988 Hawkins Stafford Education Amendments that called for improving the quality of the nation's education data. From that legislation, the National Forum for Education Statistics was begun, and from that group has followed a continuing focus on data quality issues. The Forum is made up mainly of state education agency representatives, who at times include local education agency staff in their work groups.

I have combined notes and observations from two decades of research and evaluation in public schools with the experiences from five years of reviewing and designing information systems for state and national education agencies. Often the question has been asked as to the definition of data quality and how to achieve it. The deliberations of the work groups responsible for the development of the Standards for Educational Data Collection and Reporting (SEDCAR), the ANSI ASC X-12 EDI standards for the electronic exchange of student records (SPEEDE/ExPRESS), and the national definition of dropout rates for the Common Core of Data collected by the National Center for Education Statistics have provided a unique opportunity to observe how quality is sought and defined from various perspectives. (***Getting to the Point and Counter Point of Dropout Reporting Issues***, G. Ligon, Presented at the American Educational Research Association Annual Meeting, April, 1994.) My membership on the U.S. Department of Education Evaluation Review Panel and Texas' Commissioner's Advisory Committee for Research and Evaluation has presented opportunities to relate the definitions and processes for quality data to on-going activities.

One overarching observation from these experiences is that there are multiple perspectives that determine the reality of data quality. These are generally represented by:

Decision Makers (parents, teachers, counselors, principals, school board members, tax payers, etc.)
Program Managers (principals, directors, supervisors, etc.)
General Audiences (parents, taxpayers, businesses, etc.)
Data Collectors and Providers (clerks, teachers, counselors, program managers, etc.)
Evaluators, Researchers, Analysts

Individuals may occupy more than one of these groups simultaneously.

At the risk of over simplifying, the primary perspective of each group may be described as:

> **Decision Makers:** "Do I have confidence in the data and trust in the person providing them?"
>
> **Program Managers:** "Do the data fairly represent what we have accomplished?"
>
> **General Audiences:** "Did I learn something that appears to be true and useful, or at least interesting?"
>
> **Data Collectors and Providers:** "Did the data get collected and reported completely and in a timely manner?"
>
> **Evaluators, Researchers, Analysts:** "Are the data adequate to support the analyses and the results from them?"

In this view, the burden for data quality falls to the data collectors and providers, and the evaluators, researchers, and analysts. Who else would be in a better position to monitor and judge data quality? However, in the end, the audiences (e.g., program managers, decision makers, and general audiences) give the ultimate judgment when they use, ignore, or disregard the data. Which ties in well to this paper's conclusion that *the highest level of data quality is achieved when information is valid for the use to which it is applied and when decision makers have confidence in the data and rely upon them.*

## The Pursuit of a Definition of Data Quality

Four years ago, Robert Friedman, formerly the director of the Florida Information Resource Network (FIRN) and now in a similar position for Arkansas, called and asked for references related to data quality. The issue had arisen as the new statewide education information system for Arkansas was being developed. There were few references available, none satisfactory. I began documenting anecdotes, experiences, and insights provided by individuals within the educational research, evaluation, and information systems areas to search for "truths." The resultant hierarchy is one representation of what was found.

This paper describes some of these anecdotes and experiences to illustrate the thinking of national, state, and local professionals.

Several ideas were consistently referenced by individuals concerned with data quality.

1. Accuracy

Technical staff mention reliability and accuracy. This is consistent with the published literature in the information systems area. Accuracy, accuracy, accuracy--defined as do exactly what we are told, over and over. Not all information specialists limit themselves to the mechanical aspects of accuracy; however, because they may not be content or process specialists in the areas they serve, their focus is rightfully on delivering exactly what was requested. After all, that is what the computer does for them.

> **Quality data in, quality data out.**

2. Validity

However, programmatic staff point out that data must be consistent with the construct being described (i.e., validity). If their program is aimed at delivering counseling support, then a more direct measure of outcomes than an achievement assessment is desired.

**Valid data are quality data.**

3. Investment

A key element frequently cited as basic for achieving quality is the reliance upon and use of the data by the persons responsible for collecting and reporting them. School clerks who never receive feedback or see reports using the discipline data they enter into a computer screen have little investment in the data. School clerks who enter purchasing information into an automated system that tracks accounts and balances have a double investment. They save time when the numbers add up, and they receive praise or complaints if they do not. Whoever is responsible for collecting, entering, or reporting data needs to have a natural accountability relationship with those data. The data persons should experience the consequences of the quality of the data they provide.

This may be the most important truism in this paper:

**The user of data is the best recorder of data.**

4. Certification

Typically, organizations have a set of "official" statistics that are used, regardless of their quality, for determining decisions such as funds allocation or tracking changes over time. These official statistics are needed to provide some base for planning, and the decision makers are challenged to guess how close they are.

**Organizations should certify a set of official statistics.**

5. Publication

Public reporting or widespread review is a common action cited in the evolution of an information system toward quality.

In every state that has instituted a statewide accountability system, there are stories of the poor quality of the data in the first year. Depending upon the complexity of the system and the sanctions imposed (either money or reputation), subsequent improvements in data quality were seen.

The most practical and easily achieved action for impacting data quality is:

**Publish the data.**

6. Trust

Decision makers refer to the trust and confidence they must have in both the data and the individuals providing the data.

Trust is a critical component of the working relationship between decision makers and staff within an organization. That trust must be present for data to be convincing. Consultants are used at times to provide that trust and confidence. Decision makers often do not have the time nor the expertise to analyze data. They rely upon someone else's recommendation. Data should be presented by an individual in whom the decision makers have confidence and trust.

**Trust the messenger.**

These six statements faithfully summarize the insights of professionals who have struggled with data quality within their information systems. They address processes that contribute toward achieving data quality--the dynamics influencing quality within an information system. They do not yet clearly indicate how successful the organization has been in achieving quality. To make that connection, the following hierarchy was developed.

A Hierarchy of Data Quality

A hierarchy of data quality has been designed to describe how quality develops and can be achieved. The paper details the components and levels within this hierarchy. This schema is to be regarded as fluid within an organization. Some areas of information, such as student demographics, may be more advanced than others, such as performance assessments. Some performance assessments may be more advanced than others.

The highest level of quality is achieved when data-based decisions are made with confidence. Therefore, several components of quality must be present, i.e., available data, decisions based upon those data, and confidence by the decision maker. Ultimately, quality data serve their intended purpose when the decision maker has the trust to use them with confidence. The traditional virtues of quality (e.g., reliability and validity) form the basis for that trust, but do not ensure it. Accuracy is the traditional characteristic defined within formal information systems architecture. Accuracy begs the question of whether or not the data are worthy of use.

From the observations of organizational quests for quality information systems, the concept of official data has been described. Data are official if they are designated as the data to be used for official purposes--e.g., reporting or calculation of formulas such as for funding schools and programs. At the earliest stages of information systems, the characteristic of being available is the only claim to quality that some data have. The level at the base of the hierarchy is characterized by no data being available.

Attachment A illustrates the hierarchy.

**Bad Data**

**-1.1 Invalid**

Bad data can be worse than no data at all. At least with no data, decision makers rely upon other insights or opinions they trust. With bad data, decision makers can be misled. Bad data can be right or wrong, so the actual impact on a decision's outcome may not always be negative. Bad

data can result from someone's not understanding why two numbers should not be compared or from errors and inconsistencies throughout the reporting process. The definition of bad data is that they are either:

> Poorly standardized in their definition or collection to the extent that they should be considered unusable, or
>
> Inaccurate, incorrect, unreliable.

An example of bad data occurred when a local high school failed to note that the achievement test booklets being used were in two forms. The instructions were to ensure that each student received the same form of the exam for each subtest. However, the booklets were randomly distributed each day of the testing, resulting in a mixture of subtest scores that were either accurate (if the student took the form indicated on the answer document) or chance level if the form and answer document codes were mismatched. This high school was impacted at the time by cross-town bussing that created a very diverse student population of high and low achievers. From our previous analyses, we also knew that an individual student's scores across subtests could validly range plus or minus 45 percentile points. Simple solutions to interpreting the results were not available. (***Empty Bubbles: What Test Form Did They Take?*** D. Doss and G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1985.)

Carolyn Folke, Information Systems Director for the Wisconsin Department of Education, contributed the notion that the hierarchy needed to reflect the negative influence of bad data. In her experience, decision makers who want to use data or want to support a decision they need to make are vulnerable to grasping for any and all available data--without full knowledge of their quality. The message here is look into data quality rather than assume that any available data are better than none.

**None**

**0.0 Unavailable**

Before "A Nation at Risk," before automated scheduling and grade reporting systems, and before the availability of high-speed computers, often there were no data at all related to a decision. So, this is really the starting point for the hierarchy.

When a local school district began reporting failure rates for secondary students under the Texas No Pass/No Play Law, one school board member asked for the same data for elementary students. The board member was surprised to hear that, because elementary grade reporting was not automated, there were no data available. (After a long and painful process to collect elementary grade data, the board member was not pleased to learn that very few elementary students ever receive a failing grade and that fewer fail in the lower achieving schools than fail in the higher achieving schools.) (***No Pass - No Play: Impact on Failures, Dropouts, and Course Enrollments,*** G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1988.)

When no data are available, the options are typically obvious--collect some or go ahead and make a decision based upon opinion or previous experience.

However, there is another option used by agencies involved in very large-scale data collections. The Bureau of the Census and the National Center for Education Statistics both employ decision rules to impute data in the absence of reported numbers. Missing cells in tables can be filled with imputed numbers using trends, averages, or more sophisticated prediction analyses. Decision makers may perform their own informal imputations in the absence of data.

**Available**

### 1.1 Inconsistent Forms of Measurement

Poor data come from inconsistencies in the ways in which outcomes or processes are measured. These inconsistencies arise from use of nonparallel forms, lack of standardized procedures, or basic differences in definitions. The result is data that are not comparable.

In 1991, we studied student mobility and discovered that not only did districts across the nation define mobility differently, but they also calculated their rates using different formulas. From 93 responses to our survey, we documented their rates and formulas, then applied them to the student demographics of Austin. Austin's "mobility" rate ranged from 8% to 45%, our "turbulence" rate ranged from 10% to 117%, and our "stability" rate ranged from 64% to 85%. The nation was not ready to begin comparing published mobility rates across school districts. (***Student Mobility Rates: A Moving Target,*** G. Ligon and V. Paredes, Presented at the American Educational Research Association Annual Meeting, 1992.)

A future example of this level of data quality may come from changes in the legislation specifying the nature of evaluation for Title I Programs. For years, every program reported achievement gains in normal curve equivalent units. Current legislation requires each state to establish an accountability measure and reporting system. How will performance be aggregated across states? How will gains be verified by the U.S. Department of Education as mandated?

Full time equivalents and head counts, duplicated and unduplicated counts, average daily attendance and average daily membership are all examples of how state accountability systems must align the way schools maintain their records. Who is not familiar with the "problem" of whether to count parents in a PTA meeting as one attendee each or as two if they have two students in the school?

### 1.2 Data Collected by Some at Some Times

Incomplete data are difficult to interpret.

In 1994, the *Austin American Statesman* published an article about the use of medications for ADD/ADHD students in the public schools. The headline and point of the story was that usage was much lower than had been previously reported. The person quoted was not a school district employee and the nature of some of the statistics caused further curiosity. So, I called the reporter, who said he had not talked to the District's Health Supervisor and that the facts came from a graduate student's paper. Checking with the Health Supervisor showed that only about half the schools had participated in the survey, some of those with the highest levels of use did not participate, the reporter used the entire District's membership as the denominator, and the actual usage rate was probably

at least twice what had been reported. The reporter's response: "I just reported what she told me."

### 1.3 Data Combined, Aggregated, Analyzed, Summarized

The highest level of "available data" is achieved when data are summarized in some fashion that creates interesting and useful information. At this point in the hierarchy, the data begin to take on a usefulness that can contribute to a cycle of improved quality. At this point, audiences are able to start the process of asking follow-up questions. The quality of the data becomes an issue when someone begins to use summary statistics.

One of the most dramatic responses to data I recall was when we first calculated and released the numbers and percentages of overage students, those whose age was at least one year over that of their classmates. Schools have always had students' ages in the records. Reality was that no one knew that by the time students reached grade 5 in Austin, one out of three was overage. In at least one elementary school over 60% of the fifth graders were old enough to be in middle school. (The number of elementary retention's began to fall until the rate in the 90's was about one fifth of the rate in the 80's.) (***Do We Fail Those We Fail?***, N. Schuyler and G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1984; ***Promotion or Retention,*** Southwest Educational Research Association Monograph, G. Ligon, Editor, 1991.)

When relatively unreliable data are combined, aggregated, analyzed, and summarized, a major transformation can begin. Decision makers can now apply common sense to the information. Data providers now can see consequences from the data they report. This is an important threshold for data quality. In countless conversations with information systems managers and public school evaluators, a consistent theme is that when people start to see their data reported in public and made available for decision making, they begin to focus energies on what those data mean for them and their school/program.

Texas schools began reporting financial data through PEIMS (Public Education Information Management System) in the 1980's. The first data submissions were published as tables, and for the first time it was simple to compare expenditures in specific areas across schools and districts. Immediately, a multi-year process began to bring districts more in line with the State's accounting standards and to ensure better consistency in the matching of expenditures to those categories. When districts reported no expenditures in some required categories and others reported unrealistically high amounts, the lack of data quality was evident.

***DATA BECOME INFORMATION.*** *Around this point in the hierarchy, data become information. The individual data elements are inherently less useful to decision makers than are aggregated and summarized statistics. From this point on in the hierarchy, basic data elements are joined by calculated elements that function as indicators of performance.*

**Official**

## 2.1 Periodicity Established for Collection and Reporting

Periodicity is the regularly occurring interval for the collection and reporting of data. An established periodicity is essential for longitudinal comparisons. For valid comparisons across schools, districts, and states, the same period of time must be represented in everyone's data.

> The National Center for Education Statistics (NCES) has established an annual periodicity set around October 1 as the official date for states to report their student membership. Reality is that each state has its own funding formulas and laws that determine exactly when membership is counted, and most do not conduct another count around October 1 for Federal reporting.
>
> I was called on the carpet by the superintendent once because a school board member had used different dropout rates than he was using in speeches during a bond election. He explained very directly that "Every organization has a periodicity for their official statistics." That of course is how they avoid simultaneous speeches using different statistics. After working hard with the staff to publish a calendar of our official statistics, I discovered that very few districts at the time had such a schedule. (***Periodicity of Collecting and Reporting AISD's Official Statistics***, G. Ligon et al., Austin ISD Publication Number 92.M02, November, 1992.)

## 2.2 Official Designation of Data for Decision Making

Finally, official statistics make their way into the hierarchy. The key here is that "official" does not necessarily guarantee quality. Official means that everyone agrees that these are the statistics that they will use. This is a key milestone, because this designation contributes to the priority and attention devoted to these official statistics. This in turn can contribute to on-going or future quality.

> Every year, our Management Information Department's Office of Student Records issued its student enrollment projection. The preliminary projection was ready in January for review, and a final projection for budgeting was ready by March. Here is another example of how the presence of a bond election can influence the behavior of superintendents and school board members. The superintendent gave a speech to the Chamber of Commerce using the preliminary projection. Then our office sent him the final projection. He was not happy with the increase of about 500 in the projection. He believed that created a credibility gap between the figures used in campaigning for the bonds and the budgeting process. So, the preliminary projection, for the first time in history, became the final, "official" projection. The bonds passed, the next year's enrollment was only a few students off of the "official" projection, the School Board was impressed with the accuracy of the projection, and Austin began a series of four years when all the projection formulas were useless during the oil and real estate bust of the late 80's. The next time the "official" projection was close was when a member of the school board insisted that the district cut 600 students from its projection in order to avoid having to budget resources to serve them.

> ***THE RIGHT DATA MUST BE USED.*** *At this point, the qualities of accuracy and reliability are required. Moreover, the best data are not quality data if they are not the right data for the job.*

### 2.3 Accuracy Required for Use in Decision Making

With the official designation of statistics, either by default or intent, their use increases. Now the feedback loop takes over to motivate increased accuracy. The decision makers and the persons held accountable for the numbers now require that the data be accurate.

When we began publishing six-week dropout statistics for our secondary schools, the principals started to pay attention to the numbers. They had requested such frequent status reports so the end-of-the-year numbers would not be a surprise, and so they could react if necessary before the school year was too far along. Quickly, they requested to know the names of the students that we were counting as dropouts, so verification that they had actually dropped out could be made. Having frequent reports tied directly to individual student names improved the quality of the dropout data across the schools.

> ***THE RIGHT ANALYSES MUST BE RUN.*** *The quality of data is high at this point, and the decision maker is relying upon analyses conducted using those data. The analyses must be appropriate to the question being addressed.*
>
> *A caution to data providers and audiences: There are times when data quality is questioned, but the confusing nature of the data comes from explainable anomalies rather than errors. We should not be too quick to assume errors when strange results arise. A district's overall average test score can decline even when all subgroup averages rise; students can make real gains on performance measures while falling farther behind grade level; schools can fail to gain on a state's assessment, but be improving.* (***Anomalies in Achievement Test Scores: What Goes Up Also Goes Down,*** G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1987.)

## Valid

### 3.1 Accurate Data Consistent with Definitions

Trained researchers are taught early to define operationally all terms as a control in any experiment. Every organization should establish a standard data dictionary for all of its data files. The data dictionary provides a definition, formulas for calculations, code sets, field characteristics, the periodicity for collection and reporting, and other important descriptions. Using a common data dictionary provides the organization the benefits of efficiency by avoiding redundancy in the collection of data elements. Another important benefit is the ability to share data across departmental data files. (***Periodicity™ User Guide***, Evaluation Software Publishing, Austin, Texas, 1996.)

The classic example of careless attention to definitions and formulas is *Parade Magazine*'s proclamation that an Orangeburg, South Carolina, high school reduced its dropout rate from 40% to less than 2% annually. Those of us who had been evaluating dropout-prevention programs and calculating dropout rates for a number of years became very suspicious. When newspapers around the nation printed the story that the dropout rate in West Virginia fell 30% in one year after the passage of a law denying driver's licenses to dropouts, we were again skeptical. Both these claims had a basis in real numbers, but each is an example of bad data.

The *Parade Magazine* reporter compared a four-year, longitudinal rate to a single-year rate for the Orangeburg high school. The newspaper reporter compared West Virginia's preliminary dropout count to the previous year's final dropout count. (The West Virginia state education agency later reported a change from 17.4% to about 16%.) (***Making Dropout Rates Comparable: An Analysis of Definitions and Formulas***, G. Ligon, D. Wilkinson, and B. Stewart, Presented at The American Educational Research Association Annual Meeting, 1990.)

### 3.2 Reliable Data Independent of the Collector

Reliability is achieved if the data would be the same regardless of who collected them.

What better example is available than the bias in teacher evaluations? When Texas implemented a career ladder for teachers, we had to certify those eligible based upon their annual evaluations. The school board determined that they were going to spend only the money provided by the State for career ladder bonuses, so that set the maximum number of teachers who could be placed on the career ladder. Our task was to rank all the eligible teachers and select the "best." Knowing there was likely to be rater bias, we calculated a Z score for each teacher based upon all the ratings given by each evaluator. Then the Z scores were ranked across the entire district. The adjustments based upon rater bias were so large, that near perfect ratings given by a very easy evaluator could be ranked below much lower ratings given by a very tough evaluator. The control was that the teachers' rankings within each rater's group were the same.

Everything was fine until a school board member got a call from his child's teacher. She was her school's teacher-of-the-year candidate but was ranked by her principal in the bottom half of her school, and thus left off the career ladder. The end of the story is that the school board approved enough local money to fund career ladder status for every teacher who met the minimum state requirements, and we were scorned for ever having thought we could or should adjust for the bias in the ratings. (***Adjusting for Rater Bias in Teacher Evaluations: Political and Technical Realities***, G. Ligon and J. Ellis, Presented at the American Educational Research Association Annual Meeting, 1986.)

### 3.3 Valid Data Consistent with the Construct Being Measured

The test of validity is often whether a reasonable person accountable for an outcome agrees that the data being collected represent a true measure of that outcome. Validity is the word for which every trained researcher looks. Validity assumes both accuracy and reliability. Critically, valid data are consistent with the construct being described. Another perspective on this is that valid data are those that are actually related to the decision being made.

The local school board in discussing secondary class sizes looked at the ratio of students to teachers in grades 7 through 12 and concluded that they were fairly even. Later they remembered that junior high teachers had been given a second planning period during the day, so their actual class sizes were much higher. Then they moved on to focus on the large discrepancies between class sizes within subject areas to discover that basic required English and mathematics classes can be efficiently scheduled and are large compared to electives and higher level courses. In the end, the school board members became more understanding of which data are valid for use dependent upon the questions they are asking.

## Quality

### 4.1 Comparable Data: Interpretable Beyond the Local Context

Quality is defined here beyond the psychometric and statistical concepts of reliability and validity. Quality is defined by use. Quality data are those that function to inform decision making. For this function, the first criterion is:

> Quality data must be interpretable beyond the local context. There must be a broad base of comparable data that can be used to judge the relative status of local data. We can recognize that there are some decisions that do not necessitate comparisons, but in most instances a larger context is helpful. Each time I read this criterion, I argue with it. However, it is still in the hierarchy because decisions made within the broadest context are the best informed decisions. Knowing what others are doing, how other districts are performing does not have to determine our decisions, but such knowledge ensures that we are aware of other options and other experiences.

AERA's Division H sponsors an annual publications award competition to showcase the best of the nation's evaluation reports. Each year, these can be seen in the Annual Meeting exhibit area. Educational Research Service and PDK's CEDR both disseminate these reports. The annual award recipients represent excellent examples of evaluation studies that typically provide analyses and interpretations useful beyond their local context.

> Most states and districts have struggled with defining and reporting their dropout rates. Despite the lofty goal often embraced of having 100% of our students graduate, there is still the need for comparison data to help interpret current levels of attrition. When we compared Austin's dropout rate to published rates across the nation, we found that the various formulas used by others produced a range of rates for Austin from 11% to 32%. Our best comparisons were across time, within Austin, where we had control over the process used to calculate comparable rates. (***Making Dropout Rates Comparable: An Analysis of Definitions and Formulas,*** G. Ligon, D. Wilkinson, and B. Stewart, Presented at The American Educational Research Association Annual Meeting, 1990.)

### 4.2 Data-Based Decisions Made with Confidence

The second criterion is:

> Data-based decisions must be made with confidence, at least confidence in the data. This is the ultimate criterion upon which to judge the quality of data--do the decision makers who rely upon the data have confidence in them. Assuming all the lower levels of quality criteria have been met, then the final one that makes sense is that the data are actually used with confidence.

This is a good time to remind us all that confidence alone is not sufficient. One reason the construct of a hierarchy is useful is that each subsequent level depends upon earlier levels.

> A local district's discipline reporting system had been used for years to provide indicators of the number of students and the types of incidents in which they were involved. The reports were so clear and consistent that confidence was high. As part of a program evaluation, an evaluator went to a campus to get more details and discovered that only about 60% of all discipline incidents were routinely entered into the computer file. The

others were dealt with quickly or came at a busy time. No one had ever audited a school's discipline data. On the other hand, the dropout and college-bound entries into a similar file were found to be very accurate and up-to-date.

My biases are evident in the descriptions of the levels of this hierarchy:

1. Accurate and reliable data should be a given in any information system.

2. Knowing the question being asked or the decision to be made is critical to ensuring that the right data are used and the appropriate analyses are conducted.

3. Beyond these more mechanical levels of quality, use is the goal. A claim of true quality cannot be made unless the data are useful, usable, and used.

Information systems professionals can be understood for ending their treatment of data quality somewhere in the middle of this hierarchy. For those who work at the decision-making level of an organization, more is required.

## Applying the Hierarchy to a Local School District

To illustrate whether or not the hierarchy has any relationship to a real information system, I thought back three years to our data in Austin. Attachment B is a summary of my ratings of several of the information systems from that time. These ratings range from -1.1 for the misleading data available on the computers in each school, to 4.2 for the reliable and relied upon data available on lunch and transportation programs. Yes, I rated those two areas as higher quality than assessment, in which I had invested almost 20 years. Our assessment data were excellent, but we never achieved that highest level of trust and confidence afforded lunch and transportation data. Some of that might be part of the nature of school board members' uneasiness with complex-looking test scores, or the constant tirades of detractors giving individual accounts of how test scores mislabeled their students. Assessment data will always be more challenging to control than the basic counts of who eats and who rides. But take nothing away from the lunch and bus people. They used their data, depended upon them, and ensured their quality.

## What Can an Organization Do?

A self-assessment of data quality can be conducted in each area. This can be very formal with a team approach, or very informal with a checklist kept handy for reference whenever quality issues arise.

Attachment C is a sample checklist that contains the key criteria that were identified through the development of the hierarchy. The highest level of data quality would be illustrated by a positive response to each question in the checklist.

The format recognizes that data quality will vary across areas and even across sub-areas within an area. The answers to the questions on the checklist may not be known or may be different depending upon an individual's role within the organization.

Sections *A. Statistics* and *B. Data Elements* match with levels 1.3 through 3.1 of the hierarchy. Positive ratings in these sections indicate a foundation for best practice in creating reliable, quality

data files. Section *C. Results and Interpretation* matches levels 2.2 through 3.3. Positive ratings in this section indicate that the data are being analyzed and reported for use. Section *E. Investment* fits into the hierarchy around levels 2.2 and 2.3 where the attention focused upon the data and the use of the data by the providers are key. Section *D. Confidence* represents level 4.2 where use is made of the data--with confidence.

Dealing with Error

When I read this paper just before its printing, there was a sense that the higher level nature of the hierarchy did not deal well with some of the nitty-gritty issues of data quality that are usually fretted over by information systems managers and data providers. Many of these fall into the general category of error. Error can be mistakes that result in bad data or those pesky probability statistics that keep us from ever being 100% confident in our data. *I have always been uncomfortable calling some of these problems errors when the reality is that they represent at times conscious decisions or merely differences in how data are recorded from place to place.* Error factors are divided below in two general categories.

1. Measurement Errors

Measurement errors are those imprecisions that result from our inability to be absolutely perfect in our measurements. One is the reliability of an instrument, test, or performance task (illustrated by a test-retest difference). Measurement errors can also be "intentional" as occurs when we round numbers or put values in ranges rather than use a more precise value. Sampling error limits the probability of reliable data. Measurement error is adequately dealt with in text books. Measurement error is less often adequately dealt with in practice.

At times, we lose precision by translating our data from one format to another. For example, a student's course history from one high school must be translated into the standards of another high school when the student transfers. Not only might the course content and levels not match, but the credits awarded and grading system may differ. When a California school that uses three dozen ethnicity codes for its students reports to the Office for Civil Rights, those codes are crosswalked to five categories.

2. Mistakes

These errors occur, and the challenge is to notice them, so they can be corrected if possible. Calculation errors, data entry errors, programming errors, and other human mistakes are best addressed with adequate training, monitoring, and redundancy.

Some useful techniques for detecting errors accompany the emergence of automated information systems. We now have the ability to run edit checks on data bases to determine the reasonableness of the data. Check sums can be calculated and compared to benchmark totals. Ranges of values, valid codes, and field characteristics (e.g., alphabetic, numeric, date, etc.) can be verified by the computer. Professionals always have available one of the best techniques--the use of estimating. Individuals who are goo estimators are those that are good at detecting potential errors. Use of trend data and comparable group data when available is helpful to judge the reasonableness of data.

A perspective that has become almost universal among professionals dealing with data quality issues is that when information systems became distributed throughout organizations rather than

being centralized, that the potential for errors was also distributed. The design of a distributed information system must account for data quality checks and establish responsibility for quality. The traditional notion that data processing's responsibility for accuracy begins and ends at the computer room door changed when that "door" was distributed to multiple locations through the magic of networks.

The bottom line on error is that other references have been dealing with the details of this issue for a long time. The probability issues appear to be permanent. The mistake issues have management solutions that should be employed within every organization.

## Conclusion

The hierarchy was a convenient way to think through what makes for quality data. Reality is that our information systems will not fall neatly into one of the levels of the hierarchy. In fact they may not often evolve sequentially through each level. At any point in time, their levels may shift up or down. What is useful here is that the hierarchy describes the characteristics of relatively low and relatively high levels of data quality. With the checklist and the hierarchy, an organization can begin to examine quality issues and plan improvements as needed.

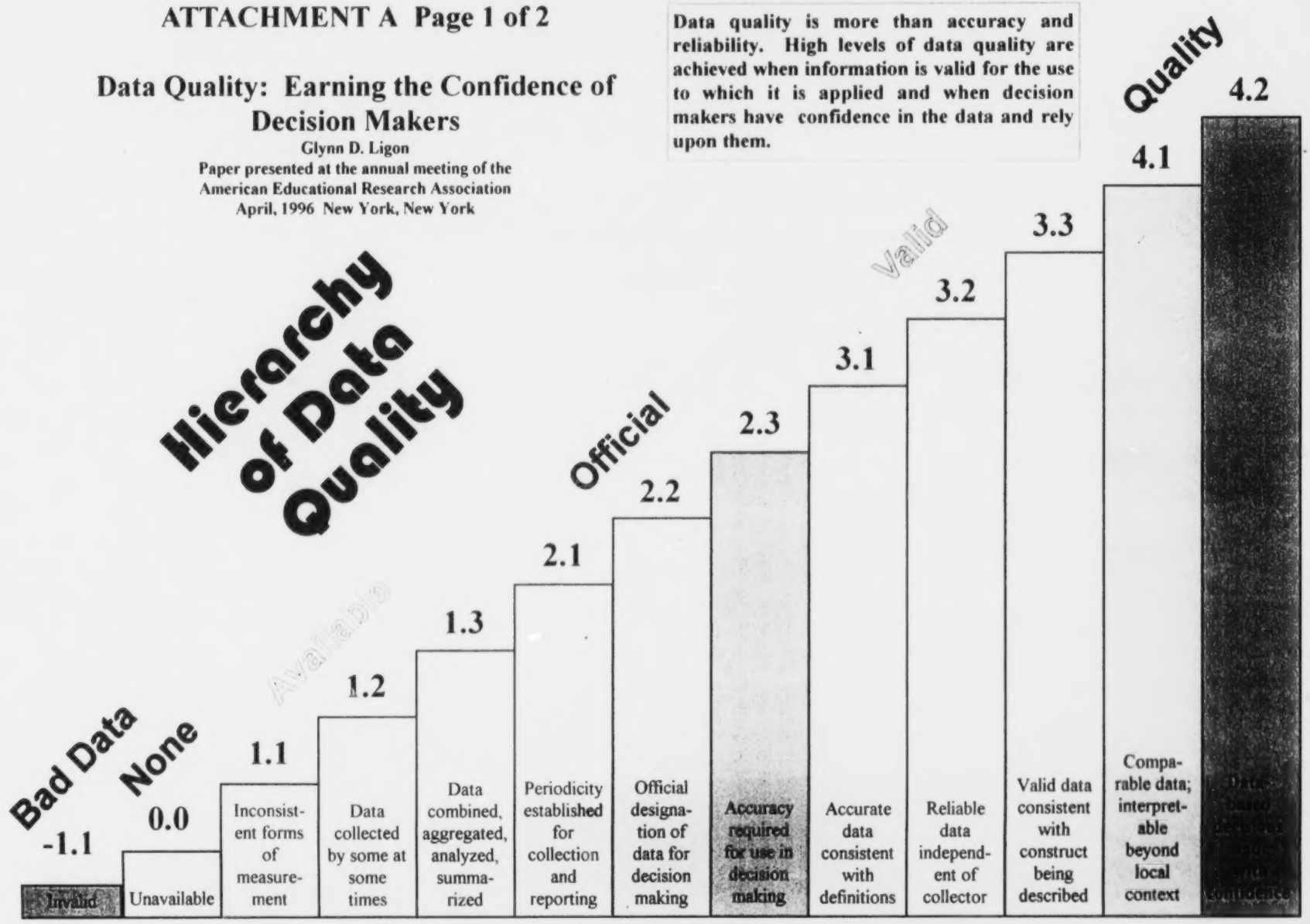
ATTACHMENT A Page 1 of 2

# Data Quality: Earning the Confidence of Decision Makers

Glynn D. Ligon

Paper presented at the annual meeting of the
American Educational Research Association
April, 1996 New York, New York

**Data quality is more than accuracy and reliability. High levels of data quality are achieved when information is valid for the use to which it is applied and when decision makers have confidence in the data and rely upon them.**




©1996 Evaluation Software Publishing, Incorporated Austin, Texas



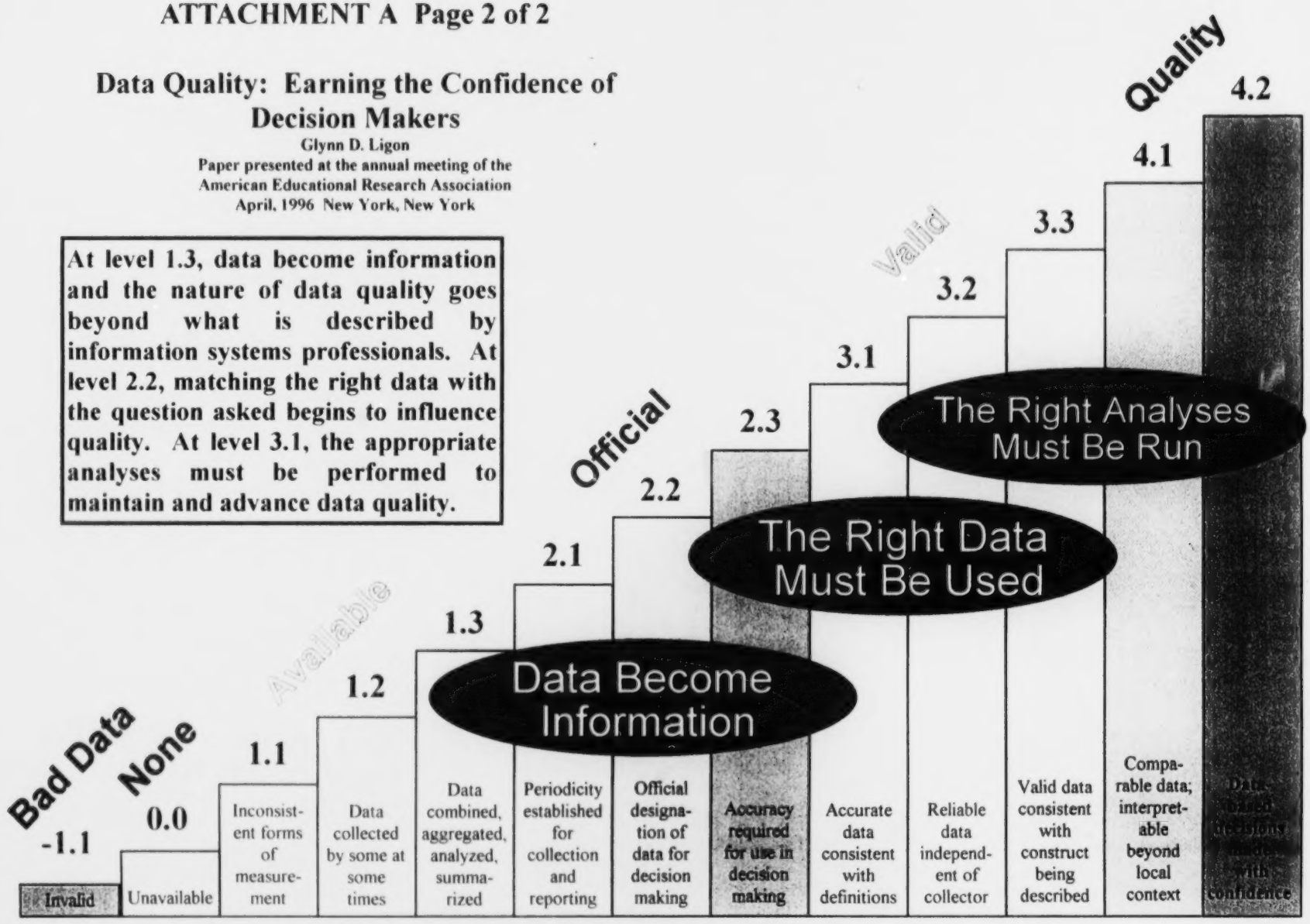
BEST COPY AVAILABLE

# Data Quality: Earning the Confidence of Decision Makers

Glynn D. Ligon

Paper presented at the annual meeting of the
American Educational Research Association
April, 1996 New York, New York

**At level 1.3, data become information and the nature of data quality goes beyond what is described by information systems professionals. At level 2.2, matching the right data with the question asked begins to influence quality. At level 3.1, the appropriate analyses must be performed to maintain and advance data quality.**






BEST COPY AVAILABLE

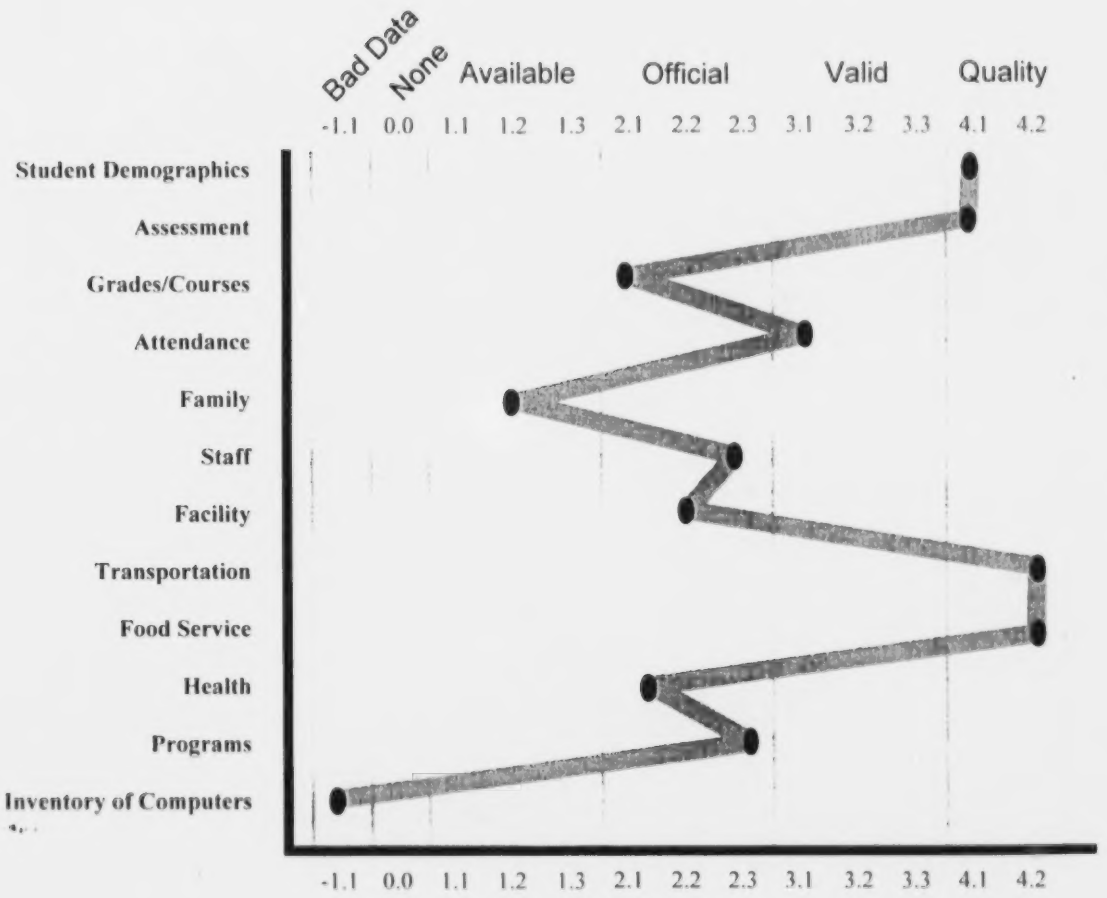
ATTACHMENT B Page 1 of 2

# Ratings of Data Quality

## Data Quality: Earning the Confidence of Decision Makers

Glynn D. Ligon

Paper presented at the annual meeting of the
American Educational Research Association
April, 1996 New York, New York




**These ratings reflect the author's opinion of the quality of data in each area during the 1992-93 school year in the Austin (TX) Public Schools. At that time, Glynn Ligon was the Executive Director of Management Information.**

# ATTACHMENT B Page 2 of 2
# Ratings of Data Quality

**COMMENTS**

**Student Demographics 4.1** Student data were automated, audited, relied upon by the schools, well defined, decision makers were just short of trusting the statistics over their own political judgment of the community. *These were some of the highest quality student demographic data in the nation.*

**Assessment 4.1** Assessment records were excellent, quality checked, testing was sometimes monitored, but not enough, longitudinal records spanned 12 years, best practice was used for preparing students and reporting results, decision makers were usually as influenced by informal information from parents, neighbors, principals, and teachers. *Longitudinal assessment data bases were excellent.*

**Grades/Courses 2.1** Course and grade information was routinely collected and reported on schedule, districtwide grading policies were in place, comparability across schools was poor (especially at elementary levels), grades and course credits could be changed by schools apart from established policy, grades were the basis for No Pass/No Play, high school and middle school grades and credits were far superior in quality to elementary grades.

**Attendance 3.1** Attendance was the basis for funding, was periodically audited, and the rules were clearly established; differences existed across schools in the reliability of the data.

**Family 1.2** Beyond very basic data elements, schools collected and maintained little family data, paper records in school offices were not consistently maintained.

**Staff 2.2** All the official data were maintained and reported; actual assignments, funding percentages for grant supported positions, stipends for extra duties, and records of professional training were inconsistent; staff evaluations were highly rater biased.

**Facility 2.2** An official description of each building was created by State law, maintaining it was difficult, bond issues prompted periodic updates, building use was poorly documented; when a decision was pending, new data were collected.

**Transportation 4.2** Records were automated for every student, scheduling and routes were computer generated, the data base was depended upon daily for operations; the District's history of bussing for integration made the data highly visible and relied upon, decision makers believed the numbers given them.

**Food Service 4.2** Money and inventory were audited and frequently reported, written applications were reviewed and audited for lunch programs, Federal requirements mandated careful accounting, decision makers liked the food service administrators and trusted them.

**Health 2.1** Immunizations were checked upon enrollment, automated records were never considered to be up-to-date or accurate, inconsistencies were often found between laws and the tracking of them within the records system.

**Programs 2.3** Students served were reasonably documented, performance reports were typically complete and submitted on time, levels of service and entry/exit dates caused some concern with analyses, decisions were more often made based upon availability of funding and allowable activities than upon evaluation findings of successful activities.

**Inventory of Computers -1.1** This rating could change often whenever the next inventory was made, counts became out-dated so fast that available data was usually misleading, older computers were not adequately differentiated from newer ones.

# Describing Data Quality
## *Checklist for Rating and Improving Data Quality*

| | |
|---|---|
| **AREA:** | **RATER:** (decision maker, program manager, collector/provider, data processor, other audiences) |
| **SUB-AREA:** | |

**RATING**

### A. STATISTICS

- ☐ 1. Are the formulas used to calculate statistics described completely in mathematical terms specifying the elements and the operations to be performed with each?
- ☐ 2. Are decision rules and standards for calculations established, e.g., precision, rounding conventions, at which steps to round, missing data options, exclusion of outliers, etc.?
- ☐ 3. Are the data elements used in the formula defined specifically to match those in the organization's data dictionary?
- ☐ 4. Are the inclusion and exclusion rules for individuals clearly defined in terms associated with each individual's record?
- ☐ 5. Are calculations conducted accurately, e.g., by competent individuals, by certified software, etc.?
- ☐ 6. Are calculated statistics accurately recorded and made available for use?
- ☐ 7. Are auditable files maintained for verification?

### B. DATA ELEMENTS

- ☐ 1. Are data elements adequately defined in the organization's data dictionary?
- ☐ 2. Are valid values, ranges, and code sets defined?
- ☐ 3. Are field characteristics, e.g., number of characters, character type, etc., defined?
- ☐ 4. Is the periodicity (the time period represented by the value of this element, e.g., point in time, one time, annual, semester, etc.) of each data element established?
- ☐ 5. Was the collection of the data element standardized and conducted accordingly?
- ☐ 6. Was the capture rate or participation rate within established target levels?
- ☐ 7. Were the data entered and processed accurately?

# Describing Data Quality
## *Checklist for Rating and Improving Data Quality*

| | | |
|---|---|---|
| **AREA:** | **RATER:** | (decision maker, program manager, collector/provider, data processor, other audiences) |
| **SUB-AREA:** | | |

**RATING**

### C. RESULTS AND INTERPRETATION

- ☐ 1. Has an official indicator or statistic been designated for use in decision making?
- ☐ 2. If so, was it used?
- ☐ 3. Were other indicators also used to provide a broader context for interpretation?
- ☐ 4. Are the results and interpretations provided consistent with the data, statistics, analyses, context, past trends, and other points of reference available? Are they reasonable?
- ☐ 5. Is the provider of the interpretations qualified for that role?
- ☐ 6. Were the questions being addressed clearly stated for each analysis conducted?
- ☐ 7. Was the appropriate statistical or analytical procedure conducted to answer the questions stated?
- ☐ 8. Were the data and statistics used in the analysis appropriate to the question being addressed?
- ☐ 9. Did the data and statistics used validly represent the construct, behavior, outcome, process, or other entity being measured?
- ☐ 10. Was information presented to describe results compared to a larger context (e.g., regional, statewide, national, international)?

### D. CONFIDENCE

- ☐ 1. Were the data provided used by decision makers?
- ☐ 2. Did the decision makers trust the data rather than seeking or relying upon other sources of information?
- ☐ 3. Were all of the decision makers' questions answered by the data provided?

### E. INVESTMENT

- ☐ 1. Do collectors of the data depend upon them in their work?
- ☐ 2. Will data be disseminated to high-use, high-interest audiences?
- ☐ 3. Are there high stakes (e.g., awards, sanctions, ratings, etc.) associated with the data?